JURNAL MATEMATIKA DAN KOMPUTER
Vol. 6. No. 1, 20 – 26, April 2003, ISSN : 1410-8518

ESTIMASI VARIANSI PADA PENARIKAN SAMPEL DUA TAHAP UNTUK DATA TIDAK LENGKAP

Sri Subanti
Jurusan Matematika F.MIPA Universitas Sebelas Maret Surakarta.

Abstract

Rasio estimation under two – phase simple random sampling is studied. A new linearisation variance estimator that makes more complete use of the sample data than a standard one is proposed. A jackknife variance estimator and its linearised version are also obtained. Unconditional and conditional repeated sampling properties of these variance estimaters are studied through simulation. Applications to 'mass' imputation under two – phase sampling and deterministie imputation for missing data are also given.

Keywords : A design consistent variance estimation; linearisation variance estimation; jackknife variance estimation.

1. PENDAHULUAN

Penarikan sampel dua tahap sering pula disebut sebagai penarikan sampel gerombol dua tahap, karena merupakan pengembangan lebih lanjut dari penarikan sampel gerombol sederhana atau yang lebih dikenal dengan penarikan sampel bergerombol. Alasan digunakannya penarikan sampel dua tahap yaitu tidak adanya kerangka penarikan sampel yang dapat mndaftarkan semua elemen atau submit dalam populasi.

Dengan demikian pada dasarnya penarikan sampel dua tahap yaitu memilih sampel acak sederhana dari chuster-chuster dan kemudian mengambil lagi sampel acak sederhana dari elemen-elemen dalam setiap chuster terpilih. Adapun teknik penarikan sampel tergantung pada perolehan informasi sebelumnya mengenai variabel pembantu x_i , estimasi rasio dan regresi membutuhkan suatu pengetahuan tentang rata-rata $\overline{X}$. Jika diinginkan untuk melapis populasi menurut x_i harus diketahui lebih dulu frekuensi distribusinya.

Jika keterangan kurang atau data yang diperoleh tidak lengkap maka untuk lebih mudahnya pada pengambilan sampel besar pertama x_i sendiri yang diukur. Tujuan sampel ini adalah untuk memberikan estimasi yang baik dari $\overline{X}$ atau frekuensi distribusi x_i . Sehingga dalam sebuah survei yang tujuannya membuat estimasi untuk beberapa variabel y_i lainnya, perhatian banyak ditujukan pada sumber-sumber dari awal walaupun ini berarti bahwa ukuran sampel dalam survei pokok mengenai y_i akan berkurang.

2. LANDASAN TEORI

Beckman dan Michael [1987] telah menyelidiki bahwa nilai harapan $\mu_1 = Ef_1\{y\}$ yang berasal dari fungsi densitas f_1 dan $\mu_2 = Ef_2\{y\}$ yang berasal dari fungsi densitas f_2. Kemudian oleh deming, d.k.k (1994) rancangan tersebut yang serupa dengan penarikan sampel rata-rata dapat menghasilkan estimasi yang tidak bias tentang μ_1 dan μ_2 .

Selanjutnya oleh Yeh, Lam dan Chi Van (1997) telah menyelidiki pula bahwa penarikan sampel sederhana untuk setiap y_1 dengan pembobot yang sama yaitu w_1 mempunyai rata-rata sampel $\overline{y}$ yang berbentuk $\sum_{i=1}^{m} w_i y_i$.

Hesterbery [1995] juga telah berhasil menyelidiki bahwa $\overline{y}$ mempunyai variansi minimu, kemudian dikembangkan oleh Fahmier , d.k.k [1985] yaitu telah memperoleh estimasi-estimasi yang tidak bias untuk $\overline{y}$ tersebut.

Kemudian Hamilton dan Mary [1995] telah mengembangkan variansi $\overline{y}$ dapat dihitung dari seluruh permutasi sehingga dapat menghasilkan juga variansi yang minimum. Oleh karena itu Krig [1998] dapat menyimpulkan bahwa variansi minimum dalam suatu sampel merupakan estimasi maksimum likelihood.

3. SAMPLING DUA TAHAP

3.1. Estimator Variansi Linier

Pada penerapan pernarikan sampel dua tahap, variabel pembantu x_i telah digunakan untuk membuat estimasi regresi dari $\overline{Y}$.

Pada sampel pertama (besar) dengan ukuran *n*' kita hanya mengukur x_i, dan pada sampel kedua sebuah subsampel acak dengan ukuran $n = vn' = \frac{n'}{k}$ dengan fraksi *v* dipilih terlebih dulu, selanjutnya diukur x_i dan y_i. Estimasi $\overline{Y}$ adalah $\overline{y}_{lr} = \overline{y} + b\left(\overline{x}' - \overline{x}\right)$ dengan $\overline{x}', \overline{x}$ merupakan rata-rata x_i pada sampel pertama dan kedua, dan *b* merupakan koefisien regresi kuadrat terkecil dari y_i terhadap x_i dihitung dari sampel kedua.

Jika tidak ada anggapan yang dibuat mengenai keberadaan sebuah regresi linier pada populasinya maka $\overline{y}_{er}$ akan menjadi bias seperti pada penarikan satu tahap. Sebuah pendekatan untuk $V\left(\overline{y}_{lr}\right)$ dapat diperoleh dengan anggapan bahwa penarikan sampel adalah acak dan $\frac{1}{n}$ serta $\frac{1}{n'}$ dapat diabaikan.

$$V\left(\overline{y}_{lr}\right) = \frac{S^2 y\left(1-\rho^2\right)}{n} + \frac{\rho^2 S^2 y}{n'} - \frac{S^2 y}{N} \qquad \text{...} \qquad (2)$$

Dalam menentukan kesalahan penarkan sampel dari $\overline{y}_{lr}$ pada penarikan sampel acak sederhana, telah ditunjukkan bahwa jika *b* dalam $\overline{y}_{lr}$ ditukar oleh koefisien $B = \frac{S_{yx}}{S_{x^2}}$ maka kesalahan dalam estimasinya sebesar $\frac{1}{\sqrt{n}}$ relatip terhadap $\overline{y}_{lr}$.

Selanjutnya diuji estimasi variansinya

$\overline{y}_{lr} = \overline{y} + B\left(\overline{x}' - \overline{x}\right)$ Misalkan $u_i = y_i - B_{x_1}$. Dalam tahap kedua, dianggap sampel besar sebagai populasi terbatas. Selanjutnya, karena sampel kecil diambil secara acak dari sampel besar maka

$E_2\left(\overline{y}_{lr}\right) = \overline{y}' : V_2\left(\overline{y}_{er}\right) = \left(\frac{1}{n} - \frac{1}{n'}\right) S_n^{1^2}$ dengan $S_n^{1^2}$ merupakan variansi *n* dalam sampel besar.

Sehingga
$$\begin{aligned} V\left(\overline{y}_{lr}\right) &= V_1\left(\overline{y}'\right) + E_1\left(\frac{1}{n} - \frac{1}{n'}\right) S_n^{1^2} \\ &= \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right) S^2\left(1-\rho^2\right) \end{aligned}$$

Karena $S_u^{1^2}$ merupakan estimasi tidak bias dari

$$S_u{}^2 = S_y{}^2\left(1-\rho^2\right)$$

sehingga diperoleh :

$$V\left(\overline{y}_{lr}\right) = \frac{S_y{}^2\left(1-\rho^2\right)}{n} + \frac{\rho^2 S_y{}^2}{n'} - \frac{S_y{}^2}{N}$$

Selanjutnya Royal [1970] menganggap bahwa populasi terbatas adalah sebuah sampel acak dari sebuah populasi yang tidak terbatas, kemudian $\overline{y}_{lr}$ menjadi model tidak bias dan hasil-hasil sampel kecil untuk variansinya dapat diperoleh, sehingga

$$E = V\left(\overline{y}_{er}\right) = \frac{\sigma^2 y\left(1-\rho^2\right)}{n} + \frac{\rho^2\sigma^2 y}{n'} - \frac{\sigma^2 y}{N} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

Persamaan (2) mempunyai bentuk sama seperti persamaan (1) kecuali pada (2) $\sigma^2 y$ dan ρ menyatakan populasi tidak terbatasnya.

Kemudian penarikan sampel dua tahap dengan regresi telah diperluas oleh Khan dan Tripathi [1967] untuk kasus *p* variabel pembantu *x* diukur dalam sampel kedua, $\overline{Y}$ dengan regresi linier berganda diestimasi denga regresi linier berganda *y* pada variabel ini. Dengan sampel kedua merupakan sebuiah sampel acak dari sampel pertama dan dengan menganggap kenormalan peubah ganda untuk *y* dan *x*, persamaan (2) untuk *p* >1 diperoleh variansi rata-rata

$$V\left(\overline{y}_{lr}\right) = \frac{S^2 y\left(1-R^2\right)}{n}\left[1 + \frac{\left(n'-n\right)}{n}\frac{\rho}{\left(n-\rho-2\right)}\right] + \frac{R^2 S^2 y}{n'} - \frac{S^2 y}{N} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

Misalkan sampel acak sederhana *s*. untuk ukuran *n*' diambil tanpa pengembalian dari populasi yang berukuran *N* dan x_i telah diobservasi untuk semua elemen $i \in s$.

Maka subsampel acak sederhana *s* untuk ukuran *n* dipilih tanpa pengembalian dari *s'* dan y_i diobservasi untuk semua $i \in s$.

Estimator rasio untuk $\overline{Y}$ adalah $\overline{y}_r = \left(\frac{\overline{y}}{\overline{x}}\right)\overline{x}' = \hat{R}\overline{x}$, dengan $\overline{y}$ dan $\overline{x}$ keduanya merupakan rata-rata untuk s dan $\overline{x}$' merupakan rata-rata untuk s'.

Estimator variansi linierisasi tak bias untuk $\overline{y}$ dinyatakan sebagai

$$V_0 = \left(\frac{1}{n} - \frac{1}{n'}\right)s^2 d + \left(\frac{1}{n'} - \frac{1}{N}\right)s^2 y \quad \ldots\ldots \quad (4)$$

dengan $s^2 d = \frac{1}{n-1}\sum_{i=s} d_i^2$; $s^2 y = \frac{1}{n-1}\sum_{i=s}\left(y_i - \overline{y}\right)^2$

$d_i = y_i - \hat{R}x_i$

Persamaan (4) variansi sampel s^2y digunakan untuk estimasi variansi S^2y.

Estimator variansi linierisasi untuk $\overline{y}_{er}$, pertama dinyatakan S^2y sebagai

$$\begin{aligned} S^2 y &= \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - R_{x_i} - R\overline{x}\right)^2 \\ &= S^2 D + 2R^S D_x + R^2 S^2{}_x \end{aligned} \quad \ldots\ldots \quad (5)$$

dengan S^2D dan S^2x merupakan variansi populasi untuk $D_i = y_i - R_{x_i}$ dan x_i, s_{D_x} merupakan variansi populasi untuk D_i dan $\boldsymbol{x}_i$.

Sedangkan $R = \frac{\overline{Y}}{\overline{x}}$.

Selanjutnya $s^2 y = s^2 d + 2\hat{R}s_{d_x} + \hat{R}^2 s^2 x \quad \ldots\ldots \quad (6)$

Berdasarkan (5) dan (6) bentuk estimator S^2y dengan $S'^2 x = \frac{1}{n'-1}\sum_{i\in s'}\left(x_i - \overline{x}'\right)^2$ menghasilkan estimator variansi linierisasi :

$$V_1 = \left(\frac{1}{n} - \frac{1}{N}\right)s^2 d + 2\left(\frac{1}{n'} - \frac{1}{N}\right)\hat{R}s_{d_x} + \left(\frac{1}{n'} - \frac{1}{N}\right)\hat{R}^2 s'^2_x \quad \ldots\ldots \quad (7)$$

Estimator ini merupakan estimator tak bias.

Dengan menggunakan persamaan (6), persamaan (4) dapat ditulus sebagai :

$$V_0 = \left(\frac{1}{n} - \frac{1}{N}\right)s^2 d + 2\left(\frac{1}{n'} - \frac{1}{N}\right)\hat{R}s_{d_x} + \left(\frac{1}{n^2} - \frac{1}{N}\right)\hat{R}^2 S_x^2 \quad \ldots\ldots \quad (8)$$

3.2. Estimator Variansi Jacknife

Pada sampling dua tahap, misalkan Y_i tidak terobservasi untuk $i \in s' - s$. Kita dapat menentukan estimator variansi $\overline{y_r}$, dengan elemen ke-j untuk setiap $j \in s'$ maka dapat ditentukan nilai $\overline{y_r}(j)$.

Selanjutnya untuk $\overline{x}$ dan $\overline{y}$ hanya jika $j \in s$ dan jika tidak $j \in s' - s$, menyebabkan $\overline{x}'$ untuk semua $j \in s'$.

Sehingga dapat didefinisikan $\overline{y_r}(j)=\left\{\overline{y}(j)\,\overline{x}(j)\right\}\overline{x}'(j)$ untuk semua $j\in s'$

dengan $\overline{x}(j)=\begin{cases}\dfrac{n\overline{x}-x_j}{n-1} & \text{untuk } j\in s\\ \overline{x} & \text{untuk } j\in s'-s\end{cases}$

$$\overline{y}(j)=\begin{cases}\dfrac{n\overline{y}-y_j}{n-1} & \text{untuk } j\in s\\ \overline{y} & \text{untuk } j\in s'-s\end{cases} \qquad (9)$$

dan $\overline{x}'(j)=\dfrac{n'\overline{x}'-x_j}{n'-1}$ untuk semua $j\in s'$

Metode Jacknife pada $\overline{y_r}(j)$ adalah

$$v_1=\frac{n'-1}{n'}\sum_{j\in s'}\left\{\overline{y_r}(j)-\overline{y_r}\right\}^2 \qquad (10)$$

Persamaan (10) merupakan estimator variansi Jacknife dengan koreksi populasi hingga $1-\dfrac{n}{N}$ dan $1-\dfrac{n'}{N}$

Untuk parameter nonlinear $\phi=g\left(\overline{Y}\right)$, estimator variansi Jacknife $\overline{y_r}(j)$ dan $\overline{y_r}$ adalah

$\hat{\theta}_r(j)=g\left\{\overline{y_r}(j)\right\}$ dan $\hat{\theta}_r=g\left(\overline{y_r}\right)$

dan

$$\overline{y_r}(j)-\overline{y_r}=\begin{cases}-\hat{R}\left(\dfrac{x_j-\overline{x}'}{n-1}\right)-\dfrac{\overline{x}'(j)}{\overline{x}(j)}\left(\dfrac{y_j-\hat{R}x_j}{n-1}\right) & \text{untuk } j\in s\\ -\hat{R}\left(\dfrac{x_j-\overline{x}'}{n'-1}\right) & \text{untuk } j\in s'-s\end{cases} \qquad (11)$$

dengan asumsi $\dfrac{\overline{x}'(j)}{\overline{x}(j)}\simeq\dfrac{\overline{x}'}{\overline{x}}$, bentuk (10) dan (11) menjadi

$$v_1\simeq\left(\frac{\overline{x}'}{\overline{x}}\right)^2\frac{s_d^{\,2}}{n}+2\left(\frac{\overline{x}'}{\overline{x}}\right)\frac{\hat{R}s_{dx}}{n'}+\frac{\hat{R}^2 s'^{\,2}_x}{n'} \qquad (12)$$

JURNAL MATEMATIKA DAN KOMPUTER
Vol. 6. No. 1, 20 – 26, April 2003, ISSN : 1410-8518

Jika $\frac{\bar{x}'}{\bar{x}} \simeq 1$ untuk n besar, maka menurut persamaan (7); persamaan (12) menjadi

$$v_2 \simeq \left(\frac{\bar{x}'}{\bar{x}}\right)^2 \left(\frac{1}{n}-\frac{1}{N}\right) s^2 d + 2\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{1}{n}-\frac{1}{N}\right)\hat{R}\, s_{dx} + \left(\frac{1}{n'}-\frac{1}{N}\right)\hat{R}^2\, s^1{}_x{}^2 \ldots\ldots\ldots\ldots\ldots \qquad (13)$$

4. KESIMPULAN

Berdasarkan uraian tersebut dapat disimpulkan bahwa, jika penarikan sampel acak sederhana dengan pengembalian maka estimator variansi linier maupun estimator variansi Jacknife merupakan estimator yang tidak bias.

DAFTAR PUSTAKA

1. Cochran, W.Gi., *Sampling Techniques,* 3rd . ed, New York, Wily, 1997.
2. Dorfman, A.H. *A note on Variance, estimation for the regression estimator in double sampling* . J. Am. Statist. Assoc.*84, 137 – 140,* 1997.
3. Raw, J.N.k. E Shaw, J., *Variance estimation under two – phase sampling with application to imputation for missing data.* Biometrika 82,2, 453-460, 1995.
4. Royall, R.M., *The prediction approach to robust variance estimation in two – stage cluster sampling* J. Am. Statist. Assoc.81,114 – 123, 1986.